

看護研究における介入効果の検証方法

多賀谷昭¹⁾

【要 旨】人を対象とする介入の効果の実験的方法による検証においては、心理的要因の制御が重要であるが、そのための標準的方法である盲検化は、看護介入の効果の検証では、通常、使用できず、ランダム化対照群比較試験 (RCT) のみを行うことが多い。これによって生じ得る問題と対策を検討し、以下の見解を得た。1) 心理的要因の影響は一般にかなり大きいと考えられる。2) プラセボ効果は本来の効果の一部として扱う必要があり、介入方法はこれを十分活用すべきである。3) ホーソン効果とピグマリオン効果の制御には、前向き観察研究か開発前後の群間比較が現実的である。4) 定常的過程への介入の効果の有無の検証では、1群前後比較が妥当な場合が多い。5) 心理的要因は常に研究者の期待に沿う方向に作用するので、その対策がないRCTの結果は信用できない。6) 効果の検証において論理的妥当性を重視する必要がある。7) 注意深いインタビューにより効果が生じる過程を明らかにできる可能性がある。

【キーワード】研究方法、看護介入、効果検証、心理的要因、RCT

1. はじめに

看護学研究における実態調査については、手法や考慮すべき条件についての知識が普及し、研究者間で意見の不一致はほとんどみられない。しかし、看護介入の効果の検証方法については、十分な合意がないように見える。この論考の目的は、その合意形成のための議論に資することである。

介入効果の検証には、一般にランダム化対照試験 randomized controlled trial (RCT) を用いることが望ましいとされているが、看護介入には特有の条件が存在するために、そのまま適用することはできない。たとえば、新薬等の効果を検証するためのRCTでは、二重盲検化を用いてプラセボや既存薬との比較が行われるが、盲検化も、盲検化に必要となるプラセボに相当する介入も、看護介入では通常不可能である。このために生じ得る問題について検討し、現実的な解決方

法を提案することを本稿の目標とする。

以下では、効果検証の対象とする介入を行う群を「実験群」と呼び、これと比較する群を「比較群」、ランダム化した比較群を「対照群」、試験研究の対象者を「被験者」と呼ぶ。

2. 実験的手法による効果検証の原理と条件

実験的手法による効果の検証では、予め選択したアウトカム変数と、十分な検出力をもつサンプルサイズを用い、予め決められた時点における評価の統計分析結果に基づいて、効果の有無や優劣の判定を行う。そのための方法として理論上もっともふさわしいのは、RCTと盲検化の併用である。

RCTでは、対照群を使用し、実験群と対照群のベースラインの条件と両群に対する(テストされる介入以外の)全ての操作を同じに揃え、予め決められ

¹⁾長野県看護大学

2013年11月8日受付

2014年3月18日受理

た時点で各群のアウトカム変数を評価する。その結果、群間に差が認められれば、その原因は介入であると判定できる。盲検化は、両群における心理的要因の影響を同等化するために使用され、心理的要因がアウトカムに影響する可能性が想定される場合には不可欠である。

RCTは、統計学的手法による効果の検証法の王道とされ、医療分野以外にも広く使われている。医療分野では単一研究として最も高いエビデンス・レベルが与えられており、メタ分析は、RCTを用いた研究結果だけを使用することが多く、看護研究でも同様である。

しかし、看護介入では、心理的要因の制御のための常套手段である盲検化が通常不可能である。看護介入では、アウトカム自体やその測定値に心理的要因が影響する場合が多く、盲検化に代わる制御方法を用いる必要があるが、後述するように、通常それはRCTとは両立しない。したがって、看護研究のエビデンスとしてRCTの結果を無条件に上位に位置づけることはできない。

3. 心理的要因とその制御

介入結果に影響する心理的要因としては、一般に3つの要素があることが知られている。効果を期待させる介入操作の形式（援助的儀式）がもつプラセボ効果、評価・観察者の期待に応えようとするホーソン効果、介入実施者が期待をもって関わることによるピグマリオン効果である。（アウトカムの測定結果には、これら以外に迎合や学習が影響する可能性がある。）

3. 1 プラセボ効果

プラセボ（偽薬や治療目的には薬効がない薬）を薬と称して投与しても、実薬と同様の効果が表れることが知られており、治療のためのプラセボ投薬も日本でも海外でも広く行われてきた。このプラセボ投薬については、効果を疑うメタ分析結果（Hróbjartsson et al., 2001; 2004）があるが、“モノ”としてのプラセボではなく、介入における“援助的儀式”としての関わりの形式によって生じる実質的な効果としての「プラセボ効果」は、肯定的に評価されており（NCCAM, 2012）、看護研究においても、プラセボ効果を考慮

に入れた効果の検証方法が論じられている（Polit et al., 2004; Sidani et al., 2011）。ただし日本では、医療行為としてのプラセボ投薬に関する倫理的観点からの批判的な調査研究（田中ら, 2011）はあるものの、看護介入に関して、「プラセボ効果」や「プラセボ介入」という用語や概念はほとんど使われておらず、看護研究法に関する小笠原ら（2012）の充実した成書においても、介入のプラセボ効果は論じられていない。「プラセボ＝偽薬」という観念連合によるマイナスイメージが、誠実さを重んじる看護の介入に関するプラセボ効果という言葉の使用を（そして、それが表す有用な概念の使用をも）妨げているのかもしれない。

3. 2 ピグマリオン効果とホーソン効果

ピグマリオン効果とホーソン効果も、迎合的評価では説明できない実際の変化をもたらす（Rosenthal et al., 1968/井上ら, 1978より引用; Rosenthal, 1994; McCarney et al., 2007）。どの薬を患者に投与しているかを医師が知ると、ピグマリオン効果によって、実薬を投与した（と医師が信じた）被験者では効果が大きく、偽薬を投与した（と医師が信じた）被験者では効果が小さくなる。また、学生の能力や努力を信じていない教師は、教育効果を上げられず、それを学生の能力不足や努力不足のせいにし、学生も教師の“負の期待”に応じて能力の発揮や努力を控えるので、ますます教育効果が低下するという悪循環が生じることは、ピグマリオン効果とホーソン効果で説明できる。

プラセボ効果を「本来の効果でない効果」とみなし、これにピグマリオン効果とホーソン効果を含めて「プラセボ効果」と呼ぶ場合があるが、この用法は誤解と混乱を招く。次節で論じるように、実証的研究のためには、3つを明確に区別する必要がある。

3. 3 心理的要因の制御

看護介入の効果に対する心理的要因の影響は、盲検化が使えないので、測定は困難であるが（Hróbjartsson, 2002）、ほかの分野の研究結果から推測することは可能である。過敏性腸症候群の針治療において、心理的要因の影響を盲検化RCTによって評価したKaptchuk et al. (2008)によれば、ホーソン効果、プラセボ効果、ピグマリオン効果は、すべて統計学的にも臨床的にも意味のある大きさであった。特に、プ

ラセボ効果とピグマリオン効果が組み合わさった場合、効果は大きく安定していた。この結果は、不安や痛みの軽減、体調の改善などを目的とした看護介入における心理的要因の影響は極めて大きいことを示唆している。

ここでいうプラセボ効果とは、介入方法自体に依存する心理的効果、すなわち、介入プログラムや特定の手法が対象者に喚起する意味付けや期待による効果を指すものとする。このように定義したプラセボ効果は、臨床応用時にも実験時と同じように作用し、また、通常の看護介入の場合、本来の効果と分離できる操作的定義は不可能なので、本来の効果の一部とみなすべきである。

介入方法、特に、最初に開発する介入方法は、プラセボ効果が存在する場合、それを十分に活用することが望ましい。そのためには、介入内容の合理性を対象者が容易に理解できるようにプログラムを構成する必要がある。明快な意味づけは、プラセボ効果を高め（Wolf, 1959）、そのことは、介入の効果全体を高めるとともに、比較対照される標準的介入方法としての適格性を高めることになる。

一方、ホーソン効果は評価を受ける場合だけに作用する。また、ピグマリオン効果は、臨床試験という状況における介入実施者の意識に依存すると考えられる。つまり、これらは、プラセボ効果と異なり、応用時に再現される保証がない。したがって、臨床に役立つ研究結果を得るためには、ホーソン効果とピグマリオン効果の制御が不可欠である。

ただし、ピグマリオン効果には、介入実施者が群間比較を意識することによって生じる効果（CPE）以外に、関係するパラダイムを意識することによって生じる効果（PPE）があると考えられる。PPEは実験的方法では本来の効果と分離できず、その意味でプラセボ効果に近い性質をもつが、文脈依存性があるので、本来の効果の一部とみなすには無理がある。PPEが存在する場合、期待通りの介入効果を得るには、関係するパラダイムが介入実施者に受け入れられることが前提となる。

ホーソン効果とピグマリオン効果の制御は、理論的にも重要な意味を持つ。これらは常に研究者の期待通

りの研究結果が出る方向に作用するので、それらの制御なしの実験は仮説の検証にならないからである。ホーソン効果とピグマリオン効果は、いずれも効果への期待の存在を前提とするので、通常は、盲検化により期待をもたらす情報を遮断することでバイアスを抑制するが、それが使えない看護介入では、両群の介入実施者・被験者とも、各介入方法が最良か同等であると信じられる状況にするほかない。この条件を満たす方法としては、次のようなものが考えられる。

- 1) 優劣が容易に判らない場合だけ群間比較による検証を行う。
- 2) 群間比較ではなく、独立した2つの自然実験デザインの観察研究とし、1群前後比較の研究結果を両者で比較する。
- 3) 新しい介入方法を開発する前に、最良と思われる既存の介入方法を比較群に実施し、その後、開発した方法を実験群に実施して結果を比較する。

このうち、1)では、心理的要因の制御自体が不要になり、2)では、比較対象がないので、介入実施者が与えられた条件下で最善の努力をする限り、ホーソン効果、ピグマリオン効果とも2つの研究間に差はないとみなすことができる。実際、よく計画された前向き観察研究の結果はRCTの結果と同じ傾向を示すという報告がある（Ioannidis et al. 2001）。外的妥当性が高いことも自然実験デザインの長所である。3)は2)と実質的には同じであるが、一つの研究として行うものである。

常套的介入方法の効果が不確かな場合の検証は1)の条件を満たしやすく、RCTの功績の例として引用される研究の大部分はこのタイプである。例えば、出産時の会陰切開に関するRCT（Harrison et al., 1984）はこの条件を満たしていたとされ（McCourt, 2005）、術前処置としての剃毛に関するRCTも初期のものは1)を満足していたと考えられる。

3. 4 アウトカムの種類と心理的要因の影響

心理的要因や迎合の影響はアウトカムの種類によって異なる。Wechsler et al. (2011) によれば、喘息患者の呼吸改善を目的とした鍼治療のプラセボ介入は、呼吸に関する計測値を改善しなかったが、主観的評価では実薬に匹敵する効果を示した。この結果は看護介

入にも当てはまるであろう。たとえば、転倒予防の介入の場合、転倒回数をアウトカムとすれば、測定に対する心理的要因の影響や迎合を無視できるが、転倒の不安の程度をアウトカムとした場合には、それらを無視できないであろう。また、肺癌患者の呼吸停止予防に関する看護介入の場合、延命率をアウトカムとした場合と、不安の程度をアウトカムとした場合では心理的要因や迎合の影響は大きく異なるであろう。生理的効果を目的とする介入では、心理尺度を効果判定のためのアウトカム変数とすることは避けるべきである。

4. 群間比較

因果関係の証明には2群比較が必要となる。1群比較では、介入前後で変化があることを証明できても、その原因を特定できない。たとえば、治療過程や妊娠過程のような非定常過程への介入の場合、自然治療や妊娠の進行による変化と介入効果を区別できない。また、アウトカム変数によっては、実質効果がなくても介入前の測定による学習や経験が介入後の測定結果を向上させる可能性がある。そうした時期的要因の影響がない定常過程への介入の場合は、1群前後比較で介入効果の有無は検証できるが、その原因が心理的要因か介入プログラムかは区別できない。

群間比較に伴う欠点や問題としては、次のようなものがある。

- 1) 1群前後比較に比べて、多くの被験者を必要とする。
- 2) 不公平に対する倫理的配慮を要し、比較群の被験者を確保しにくい。

このうち1)については、後述するランダム化に伴う制約のほか、介入条件の同等化に伴う困難が存在し、これには、介入実施者の割り付けや介入方法の統一などが含まれる。2)については、実験デザインとその意義を被験者に理解してもらいにくいことが問題である。特に、RCTの場合、インフォームド・コンセントの署名があっても、それが実質的理解に基づく同意によるものかどうか疑わしいことが少なくない (Appelbaum et al., 1982)。

比較群使用の代替法としては、次のようなものが考えられる。

- 1) 実験研究ではなく、比較に適する対象の前向き観察研究を行う。
- 2) 1群のみを用いて前後比較を行う。(ただし、定常過程への介入で、測定に学習や経験が影響しないアウトカムを用いることが条件である。)
- 3) ベースラインの同等化の有無にかかわらず、介入前後の差を比較する。

このうち1)は、形式上、群間比較ではなく独立した観察研究になり、これとは独立して行う実験群と結果を比較することになるが、倫理的な問題が生じにくく、後述する心理的要因の制御にも有利である。3)は常に可能で、RCTやその他の手法と併用できる。この方法では交互作用がチェックできないが、一般に交互作用は主効果に比べてかなり小さいので、ベースラインの差が比較的小さければ、無視できる。

5. ランダム化

ランダム化は、実験群と対照群のベースライン条件の同等化に用いる。RCTがgold standardとみなされるのは、ランダム化によって未知の条件を含むすべての条件を同等化できるという理論上のメリットのためである。しかし、次のようなデメリットがメリットを上回る可能性もある。

- 1) 同時に2つの群を扱う必要があり、施設内での実施が難しい。
- 2) RCTの研究デザインは被験者に理解されにくく、真のインフォームド・コンセントが成立しにくい。
- 3) ランダム化した後の脱落者が多く、同等性が損なわれる。
- 4) 被験者の心理的特性に偏りが生じるために、結果が一般化しにくい。

このうち、2)、3)、4)は、ランダム化がもつ一種の不自然さが人の気持ちや判断をかき乱す結果である。それは、倫理的な問題であると同時に、研究結果の価値にも悪影響を与える。これらのことから、研究の実施においても、研究結果の信頼度の評価においても、RCTを絶対視すべきではないと考えられる。特に2)の問題は深刻で、Appelbaum et al. (1982)によれば、無介入対照群に割り当てられても、治療を受けられると勘違いしていた被験者が多かった。また、

RCTでは 4)以外にも、選択条件の厳しさなど、標本代表性を損なう要素があり、外的妥当性は一般に低い (Rothwell, 2005)。

上のような問題が少ない代替法として次のようなものが考えられよう。

- 1) 介入前後の差だけを両群で比較する。
- 2) 時期的要因が無視できる場合は、時期を変えて群ごとに偶然に任せたサンプリングを行うことにより、ランダム化に近い同等化ができる。
- 3) 影響要因が既知であれば、同等化できなくても、影響要因を測定して補正することが可能である。

このうち1)は、前述のように、ベースラインの同等化に関する要求水準を緩和する。2)と1)の組み合わせは、現実的な解決策であり、さらに3)も併用可能である。

6. アウトカム変数とサンプルサイズの設定

効果の評価には、適切なアウトカム変数とサンプルサイズを採用する必要がある。アウトカム変数もサンプルサイズも、事前に選定する必要がある。結果を知ってから変更すると、その評価結果は信用できない。報告では、推定される効果（またはその差）の大きさとサンプルサイズから算出される検出力を記載する必要がある。なお、2群比較の場合、一方の群だけサンプルサイズを増やしても検出力はわずかしかならない。たとえば、等分散なら25対1000よりも50対50の方が検出力は高い。

6. 1 アウトカム変数の選択

アウトカム変数には効果の有無等を示す二分変量と効果の程度を示す連続変量の二種類がある。二分変量は意味が明確で扱いやすいが、検出力が低く、連続変量は検出力が高いが、効果の大きさの実用的な価値は判断しにくい。したがって、効果の判定には連続変量、実用性の評価には二分変量を使用するのが合理的である。しかし、変量の種類にかかわらず重要なのは、心理的要因に対する頑健性や安定性であり、特にRCTではそれが不可欠である。

6. 2 サンプルサイズの算出

サンプルサイズは小さすぎても大きすぎても協力者の負担や研究費用を浪費することになるので、適切な

サンプルサイズの設定は、倫理的にも経済的にも非常に重要である。先にあげたHulley et al. (2001) の日本語版 (木原ら, 2004) では、320頁中41頁がサンプルサイズの見積もりに関する記述にあてられている。検出力分析のためのソフトウェアは、インターネット上に公開されている。例えば、ドイツのデュッセルドルフ大学からG*Power3という無料の分析ツールが提供されている。

検出力分析では、アウトカム変数について、文献や経験、その他の情報から介入の効果の大きさeffect sizeを推測して、その検出にかなりの確率で成功するためのサンプルサイズを算定する。通常は有意水準0.05、検出力80%~90%に設定して計算する。ただし、効果に差がないことを示すことが目的の場合は、検出力を95%以上に設定する。

検出力を100%に近づければ、必要なサンプルサイズは急激に上昇するので、経済的に研究が成り立たなくなる。したがって、量的研究を行う者は、10%から20%程度は有意差の検出に失敗する可能性を受け入れなければならない。これは決して低い可能性ではないが、複数のアウトカム変数を使用すれば、実際に差があるのに有意差が全く検出できないという事態は、まず避けられる。これによって、一部のアウトカム変数だけ有意になる確率も高くなるが、複数のアウトカム変数が質的に異なっていれば、結果の解釈に窮することはない。都合の良い結果だけを報告することは研究倫理に反する。

看護介入では、対象者や協力施設の負担を伴う場合が多く、倫理的にも経済的にも、検出力分析が必要なことは明らかであるが、実際には行われずに検出力が低すぎる研究が多くみられる。検出力が低すぎる研究は、有意差が検出できない場合は結果を公表できずに研究協力者の善意を無にし、運よく有意差が検出された結果だけを報告して他の研究者の判断を誤らせることになる。介入研究に事前登録が求められる理由の1つは、この公表バイアスを防止するためである。

実際には、効果の大きさは通常何らかの方法で推定できるはずで、全く推定できないとすれば、準備不足である可能性が大きい。それでも情報が不足する場合は、臨床的に意味がある効果の下限を用いる。どう工

夫しても有意差の検出（あるいは差がないという証明）に必要なサンプルサイズが非現実的な大きくなる場合は、実施しても意味のある結果が得られる見込みが小さいので、研究デザインを変更する必要がある。

アウトカム変数として二分変量を採用することは、ほとんど常に可能である。例えば、既存の介入方法は被験者の40%に改善をもたらし、開発した方法は70%程度に改善をもたらすと推定される場合、検出力を90%にするには、各群62人の被験者が必要になる。効果がある被験者の割合が推定できない場合は、臨床上意味があると考えられる最低限の割合を使用すればよい。上の例で60%の被験者に改善を認めれば開発した方法に意味があると考えられるならば、各群139人としておく。研究を行おうとする事象について、この程度の見積もりができないことは、まずあり得ないので、検出力分析は常に可能だと考えて差支えない。

ただし、臨床上意味があると考えられる最低限の割合を使用する場合は、サンプルサイズが過大に見積もられる可能性が大きく、効果を証明する十分なデータが得られた段階で中止する方が、倫理的にも経済的にも望ましい。中間の有意差検定は多重仮説検定になるので、オプライエン・フレミング法（O'Brien et al., 1979）などによる有意水準の補正を行う。また、中間の評価は、介入に携わらない実験管理者が行い、実験継続中は評価結果を介入実施者に秘匿する必要がある。

7. 現実的な研究デザイン

以上のことから、心理的要因の影響を考慮した現実的な研究デザインは次のようになる。

なお、ここで特に触れない事項については、一般の試験研究と同じとする。たとえば、季節性や実施施設の特性、非定常的過程における被験者のステージなど、ベースラインの条件をできるだけ同等化し、また、効果が継時的に変化する場合は、効果を測定する回数と時点を決定しておく必要がある。

7. 1 効果の優劣や違いの検証

既存の介入方法が存在する場合には、その効果と新規開発の介入方法の効果の優劣や性質の違いを比較することが望ましい。それによって、無介入の比較群を

用いることによる倫理的な問題を回避するとともに、

（2つの方法の心理的要因が等しいとみなせば、）介入方法自体の実質的な差を明らかにすることが可能になる。ただし、効果の差の検証は、効果の有無の検証に比べはるかに多くの被験者を必要とするので、次の7.2の方法によって、新規開発の介入方法に効果があることを確認した上で実施すべきである。

検証には群間比較が必要で、形式上、次の2つがあり得る。

- 1) 1つの実験的研究の中の群間比較として実施する。
- 2) 独立した2つの観察研究として実施し、その結果を比較する。

RCTを実施するには1)を採用する必要があるが、心理的要因を制御できることがその前提となる。テストする介入方法を開発する前に比較群に既存の介入を実施する場合や、優劣が推測しにくい場合は、この研究デザインを採用することが可能である。一方、2)は、介入実施者も対象者も比較を意識しないので、ホーソン効果やピグマリオン効果の影響を受けにくいという利点がある。いずれにしても、RCTを使用しない場合は、介入前後の差を比較に用いるなど、ベースラインの同等化の不完全さへの対策を行う必要がある。

7. 2 効果の有無の検証

効果の有無を検証するための研究デザインは、介入の対象となる過程が定常的か否かによって異なる。特別な場合以外は、無介入比較群を用いる必要があるが、その場合、好ましいと考えられる介入をあえて控えることで倫理的な問題が生じる可能性がある。その代償として研究の後で行う介入は、定常的過程では可能であるが、非定常的過程では不可能である。介入方法の開発前に行った実態調査の結果を比較に用いればこの問題を回避できる。

いずれにしても、通常、看護介入の実質効果と心理的要因の影響を分離することはできないので、実質的效果の存在を主張するには、心理的要因の影響を受けにくいアウトカム変数を使用するとともに、介入プログラムの内容に理論的な妥当性があることを示すことが重要になる。また、観察やインタビューも実質的效果の存在の証明に役立つかもしれない。

7. 2. 1 定常的過程への介入の効果の有無

慢性的症状や睡眠など、定常的過程や日常的に繰り返される非定常的過程に対する看護介入の効果の有無は、アウトカム変数の値が介入以外の影響を受けないならば、アウトカム変数の値を介入前後で比較するだけで検証できる。この場合、被験者が少なく、かつ、検出力が大きくなる点が好都合であり、また、無介入による倫理的問題も生じない。

記銘力テストの場合のような学習効果や血圧などのような平均回帰が存在する場合は、無介入の比較群を使用して学習効果や平均回帰によるアウトカム変数の変化を測定する必要がある。

7. 2. 2 非定常的過程への介入の効果の有無

治癒過程や妊娠・出産や退院など、日常的には繰り返されない非定常的過程に対する看護介入の場合は、群間比較が不可欠である。被験者のステージを描いた上で、比較群には全く介入を行わずに（非定常的過程における）変化を追跡し、実験群の変化と比較する。アウトカム変数の測定の回数と時期は、予備試験によって慎重に決める必要がある。

ただし、効果の持続期間が比較的短く、その期間内で局所的に定常的過程とみなすことが可能であれば、上述の定常的過程への介入（7.2.1）の場合に準じて扱うことができる。

7. 3 検出力分析

妥当で敏感なアウトカム変数を選定し、様々の知識や経験を活かして効果の大きさを推定し、サンプルサイズを算出する。推定が困難な場合はアウトカム変数に二分変量（binary outcome variable）を用いて効果の出現率を推定する。複数のアウトカム変数を用いて効果を多面的に評価することができるが、同じ側面を複数の変数で測定する場合は、そのうちどの変数を判定に用いるかを予め決めておく必要がある。

連続量（continuous outcome variable）を用いる方が、一般に検出力が高くなるが、ビジュアル・アナログ・スケール（VAS）を採用する場合には注意が必要である。効果自体への心理的要因の影響の有無にかかわらず、VASによる測定値が心理的要因の影響を受けることは明らかで、使用する場合は心理的要因の制御が不可欠である。

算出したサンプルサイズがどのアウトカム変数につ

いても大き過ぎて、実現不可能な場合は、研究を実施しても有意差を検出できる見込みがないので、研究デザインを変更する必要がある。それは、評価方法の変更だけで済むかもしれないが、実験デザインの大幅な変更が必要かもしれないので、検出力分析は、研究計画作成の早い段階で実行すべきである。

7. 4 実験の管理

必要に応じて実験を中止するために介入結果を監視する場合は、介入を実施しないで監視を行う実験管理者を置く。途中で有意差検定を行う場合は、多重仮説検定になるので、オブライエン・フレミング法などで有意水準の補正を行う。

7. 5 インタビューによる効果の検証

量的方法で有意な効果が認められる場合や、理論的に効果が確実視される場合、インタビューによって効果を生じるプロセスを明らかにすることにより、効果が介入プログラム自体に起因することを確認できる可能性がある。ただし、心理的要因の影響を避けるには、介入を受けて初めて表現できるような語りに着目する必要がある。介入の効果の有無や程度を直接尋ねてもあまり意味がない。

8. 効果検証の限界

8. 1 実験的検証結果の情報としての価値

実験的手法による効果検証の重要度は、看護介入プログラム開発の場合と新薬開発の場合とは異なる。多くの看護介入では、（心理的效果を含めた介入全体として）効果があることは、ほぼ確実である。ベイズ統計学の用語で表現すれば、効果があるという事前確率がきわめて高い。したがって、それを証明できたとしても、総合的判断における情報としての重要度は相対的に低い。つまり、全く新しい介入プログラムの開発では、確認のために効果の有無を検証することは研究の手続きとして不可欠ではあるが、情報としては、事前確率を高める介入プログラムの内容に主な価値がある。それに対して、既存の介入方法の効果検証や、新旧の介入プログラムの間で効果を比較する場合は、実験結果が重要な情報となる。研究によって得られる情報の価値が文脈に依存するだけでなく、後述するように、実験結果自体も文脈に依存する。

8. 2 心理的要因によるバイアス

上述したように、看護介入の効果に心理的要因が影響する場合は、盲検化なしで実施されたRCTの結果の信頼性には疑問があり、エビデンスの序列を再検討すべきである。特に、そのバイアスが、常に研究者が期待する方向に作用することは大きな問題である。心理的要因の影響が想定される場合、メタ分析に盲検化を伴わないRCTの結果を使用すると、いくら多くの研究結果を用いても分析結果は信用できない。結果の誤差に一定の方向性がない点で、観察研究の方がRCTに優る。

8. 3 実験的検証の文脈依存性

8. 3. 1 必要なエビデンス・レベルの文脈依存性

検証に要求されるエビデンス・レベルは文脈に依存する。論理的妥当性によって因果関係が十分説明でき、特に危険がない場合にレベルの高いエビデンスを求めることは合理性を欠く。さらに、検証の結果自体も、次の例で示すように、歴史的な文脈に依存する。

実証的研究の重要性を示す例として、手術前の剃毛が術後感染予防に役立たないことの証明がしばしば引用され（例えばDiCenso et al. (1998)）、近年も多くの実験的研究が行われている（Jose et al., 2013）。しかし、この問題に関して、実験による確認は手続として必須ではあるが、その意義は限られている。

皮膚と体毛の解剖学的・生理学的・細菌学的特徴から、剃毛が皮膚の身体保護機能と皮膚自体の健康を損なうこと、また、体毛は皮膚より消毒しやすく、剃毛には作業を容易にする以外の意味がないことは、実験を待たずとも明らかである。さらに、下腹部や頭部の剃毛が患者の精神的健康を害することも自明である。したがって、作業の難易に影響がなければ、術後感染リスクは剃毛した場合の方が大きいことには理論的な根拠がある。

この検証では、一応、身体各部で剃毛の有無と消毒後の皮膚の細菌学的状態との関係を確認した後、手術の種類ごとに体毛の存在が作業に影響する程度を査定し、どの程度まで剃毛なしの手術が許容できるかを臨床試験によって確認することになる。近年、頭部の手術においてさえ、術前剃毛が術後感染率を減少させないことが示されている（Sebastian, 2012;

Broekman et al., 2011）。したがって、これより作業が容易な手術における実験的検証は理論的には不要である。

8. 3. 2 検証結果の文脈依存性

臨床試験では、ホーソン効果や群間比較によって生じるCPEは観察研究の方法で制御できるが、パラダイムによるPPEは制御できない。たとえば、剃毛しない場合、切開や縫合の作業は体毛を避けて行うため、より多くの注意と時間を要する。この作業中、実験の意味を理解している医師は清潔や手術部位の状態に剃毛した場合より多くの注意を払うので、実験結果は常に研究者の期待に沿う方向にずれるであろう。したがって、臨床試験で検証できる仮説は、「剃毛なしでも、他の条件が同じなら、術後感染のリスクは変わらないかむしろ下がる」という命題ではなく、「剃毛なしでも、改革精神をもって十分注意して手術すれば、術後感染のリスクは変わらないかむしろ下がる」という下線部の条件がついた命題である。

このことが示すように、PPEの影響を受ける介入の効果の実証的検証は、特定のパラダイムを前提とする条件がついた仮説しか扱えないので、検証結果は前提となるパラダイムの文脈でしか効力がなく、また、その条件のあいまいさに起因する誤差はメタ分析では克服できない。さらに、ほぼ定説となった仮説の検証では、公表バイアスも大きくなりやすい。（定説と異なる結果は公表されにくい。）

EBMのパラダイムとは無関係に、剃毛しない方がよいことは信用できる。しかし、その根拠は理論的妥当性と精神的なダメージを回避できるという明らかなメリットおよび不確実性ゆえにPPEが働かなかったフロンティアにおける臨床試験の結果にある。その後確認のために行われた多数のRCTやそのメタ分析の結果は、勇気づけられるものではあるが、科学的証拠としての効力は疑わしい。

8. 3. 3 研究結果の応用における文脈依存性

介入方法の臨床試験の結果はほぼ常に特定のパラダイムを前提とするので、それを臨床に導入する際、期待した効果を得るには、臨床試験と同じ文脈が前提となる。さらに、実験的研究の結果の外部妥当性が低いことにも注意する必要がある。例えば、手術内容や患

者の状態によって術後感染に対する剃毛の影響は変化するが、研究では複雑な条件の事例は除外されるので、RCTやメタ分析の結果を絶対視することはできない。臨床への応用において期待した効果を得るとともに機械的適用による危険を避けるには、手法の伝達だけでなく、その基盤にあるパラダイムと根拠となる理論的妥当性の理解を伴う必要がある。手術室から剃刀を強制的に排除するだけでは目的は達成できない。特に、途上国における保健医療の状況を評価するには、EBMやEBNにもパラダイム依存性があることに注意する必要がある。

8. 4 実験的検証の位置づけ

研究においても実践においても、本質的に重要なものは、科学的で批判的な思考方法である。それは、実験的研究から応用へという直線的過程ではなく、循環的な過程である (McCourt, 2005)。すなわち、常識を見直し、問題に気付く、分析し、対策を立て、その結果を検証する、という過程を繰り返す必要があり、検証はその一部にすぎない。また、既存の実験的検証の結果の総合的検討は、機械的なメタ分析ではなく、文脈を考慮したシステマティック・レビューによる必要がある。

研究方法としても、実践のための根拠としても、RCTが最善とは限らない。雑誌Evidence Based Nursingの編集者 (DiCenso et al., 1998) が主張するようなRCTの絶対視は、常套的介入方法の改善のための政治的正当性を有するが、上述のように心理的要因が考慮されておらず、ナイーブすぎる見方である。

9. まとめ

臨床試験の効果検証の一般的方法とその基礎理論を参考にして看護介入の効果判定方法を検討した結果、次のようなことがわかった。

- 1) プラセボ効果を実質的效果と分離して測定することは不可能で、本来の効果の一部とみなす必要がある。
- 2) 心理的要因の影響を受ける介入方法は、合理性が明らかで、プラセボ効果を十分活用できることが望ましい。
- 3) 心理的要因の影響を受けにくいのは、観察研究か

開発前後群間比較であり、倫理的な問題も回避できる。

- 4) 介入プログラムに実質効果があり得ることをインタビューによって証明するには、心理的要因の影響を避けるスキルと分析方法が必要である。
- 5) 妥当で、かつ、可能な限り心理的要因の影響を受けないアウトカム変数を選定し、研究計画の早期に検出力分析を行う必要がある。
- 6) 看護介入に関するRCTの結果は、研究者の期待に沿う方向にずれる可能性が大きく、エビデンス・レベルの序列の有用性を見直す必要がある。
- 7) 心理的要因の影響が無視できない場合、メタ分析には合理性がなく、文脈を十分考慮したシステマティック・レビューが必要である。

謝 辞

この論考の執筆は、紀要の編集や倫理審査に携わった経験がきっかけになった。多くの重要な問題に気づかせてくださった投稿者、申請者、委員の皆様に感謝申し上げます。

文 献

- Appelbaum P.S., Roth L.H., Lidz C. (1982): The therapeutic misconception: informed consent in psychiatric research, *Int J Law Psychiatry*, 5 (3-4), 319-29.
- Broekman M.L., van Beijnum J., Peul W.C., et al. (2011): Neurosurgery and shaving: what's the evidence? , A review. *Journal of neurosurgery*, 115(4), 670-678.
- DiCenso A., Cullum N., Ciliska D. (1998): Implementing evidence-based nursing: some misconceptions, *Evidence Based Nursing*, 1(2), 38-39.
- Harrison R.F., Brennan M., North P M., et al. (1984): Is routine episiotomy necessary? , *Br Med J (Clin Res Ed)*, 288(6435), 1971-1975.
- Hróbjartsson A., Gøtzsche P.C. (2001): Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment, *New Engl J Med.*, 344(21), 594-1602.

- Hróbjartsson A., Gøtzsche P. C. (2004): Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment, *Journal of Internal Medicine*, 256, 91-100.
- Hróbjartsson A. (2002): What are the main methodological problems in the estimation of placebo effects? *Journal of Clinical Epidemiology*, 55(5), 430-435.
- Hulley S.B., Cummings S.R., Browner W.S., et al. (2001): *Designing Clinical Research: An Epidemiologic Approach*, Lippincott Williams & Wilkins, Philadelphia, P.A. / 木原雅子, 木原正博 訳 (2004) : 医学的研究のデザイン: 研究の質を高める疫学的アプローチ第2版, メディカルサイエンスインターナショナル, 東京.
- 井上健治, 大沢啓子, 亀谷秀樹, 佐々木正宏, 渡辺孝憲 (1978), 教師の期待効果に関する研究, 東京大学教育学部紀要, 17, 59-76.
- Ioannidis J.P., Haidich A.B., Pappa M., et al. (2001): Comparison of evidence of treatment effects in randomized and nonrandomized studies, *J Am Med Assoc*, 286(7), 821-30.
- Jose B., Dignon A. (2013): Is there a relationship between preoperative shaving (hair removal) and surgical site infection?. *Journal of Perioperative Practice*, 23(1), 22-25.
- Kaptchuk T.J., Kelley J.M., Conboy L.A., et al. (2008): Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome, *Bmj*, 336(7651), 999-1003.
- NCCAM=National Center for Complementary and Alternative Medicine in National Institute of Health (2012.12.12): "Placebo Effect", <http://nccam.nih.gov/health/placebo>.
- McCarney R., Warner J., Iliffe S., van Haselen R., et al. (2007): The Hawthorne effect: a randomised, controlled trial, *BMC Medical Research Methodology*, 7(1), 30.
- McCourt C. (2005): Research and theory for nursing and midwifery: Rethinking the nature of evidence, *Worldviews on Evidence-Based Nursing*, 2(2), 75-83.
- O'Brien P.C., Fleming T.R. (1979): A multiple testing procedure for clinical trials, *Biometrics*, 35, 549-612.
- 小笠原知枝, 松本光子編 (2012) : これからの看護研究—基礎と応用—, ヌーベルヒロカワ, 東京.
- Polit D.F., Beck C.T. (2004): *Nursing Research: Principles and Methods*, Lippincott Williams & Wilkins, PA.
- Rosenthal R. (1994): Interpersonal expectancy effects: A 30-year perspective, *Current Directions in Psychological Science*, 3(6), 176-179.
- Rosenthal R., Jacobson L. (1968): *Pygmalion in the Classroom*, Holt Rinehart & Winston, New York.
- Rothwell P.M. (2005): External validity of randomised controlled trials: to whom do the results of this trial apply?, *Lancet*, 365 (9453), 82-93.
- Sebastian S. (2012): Does preoperative scalp shaving result in fewer postoperative wound infections when compared with no scalp shaving? A systematic review, *Journal of Neuroscience Nursing*, 44(3), 149-156.
- Sidani S., Branden C. J. (2011): *Design, Evaluation, and Translation of Nursing Interventions*, Wiley, Hoboken, NJ.
- 田中美穂, 小松明 (2011) : 臨床における看護師のプラセボ与薬の実態に関する全国調査, 日本看護倫理学会誌, 31(2), 36-46.
- Wechsler M.E., Kelley J.M., Boyd I.O., et al. (2011): Active albuterol or placebo, sham acupuncture, or no intervention in asthma, *New England Journal of Medicine*, 365(2), 119-126.
- Wolf S. (1959): The pharmacology of placebos, *Pharmacological Reviews*, 11, 689-704.

【Special Contribution】

Problems in trials of nursing interventions and their realistic solutions

Akira TAGAYA¹⁾

¹⁾Nagano College of Nursing

【Abstract】 Since blinding and randomization are usually unrealistic by ethical and practical reasons in testing effects of nursing interventions, alternative solutions are required for dealing with psychological biases and inequality of baseline conditions between groups. The following findings were obtained. 1) The placebo effect, Hawthorne effect, and Pygmalion effect can be clinically very significant. 2) The placebo effect must be regarded as a part of the net effect of intervention because, unlike other two components of psychological effects, it cannot be controlled in quantitative studies, hence a completely new intervention program should be persuasive to the participants about its effectiveness to maximize its placebo effects, which will qualify the program as a standard intervention for reference. 3) Observational research or before-and-after-development two-group comparison are realistic solutions for equalizing the Hawthorn and Pygmalion effects between groups. 4) Use of one-group before-after comparison is often justifiable for interventions in non-transitional processes. 5) Results of RCTs are unreliable without controlling psychological effects because resulting biases are always in favor of the researchers' expectations; hence evidence levels of RCTs and observations require reexamination. 6) Reasoning is sometimes more appropriate for examination of effects than an experimental trial. 7) Interviews can reveal the process of effects but cannot prove existence of the effects.

【Keywords】 research methodology, nursing intervention, test of effect, psychological effects, RCT

多賀谷昭

〒399-4117

長野県看護大学

Tel: 0265-81-5100 Fax: 0265-81-1256

Akira TAGAYA

NaganoPrefecture

Nagano College of Nursing

1694Akaho,Komagane,Nagano,399-4117JAPAN

TEL: +81-265-81-5100 FAX: +81-265-81-1256

E-Mail:tagaya@nagano-nurs.ac.jp